

# Breaking the $\ell_1$ Recovery Thresholds with Reweighted $\ell_1$ Optimization

Weiyu Xu  
Caltech EE  
weiyu@systems.caltech.edu

M. Amin Khajehnejad  
Caltech EE  
amin@caltech.edu

A. Salman Avestimehr  
Caltech CMI  
avestim@caltech.edu

Babak Hassibi  
Caltech EE  
hassibi@systems.caltech.edu

**Abstract**—It is now well understood that  $\ell_1$  minimization algorithm is able to recover sparse signals from incomplete measurements [2], [1], [3] and sharp recoverable sparsity thresholds have also been obtained for the  $\ell_1$  minimization algorithm. In this paper, we investigate a new iterative reweighted  $\ell_1$  minimization algorithm and showed that the new algorithm can increase the sparsity recovery threshold of  $\ell_1$  minimization when decoding signals from relevant distributions. Interestingly, we observed that the recovery threshold performance of the new algorithm depends on the behavior, more specifically the derivatives, of the signal amplitude probability distribution at the origin.

**Index Terms:** compressed sensing, basis pursuit, Grassmann angle, reweighted  $\ell_1$  minimization, random linear subspaces

## I. INTRODUCTION

In this paper we are interested in compressed sensing problems. Namely, we would like to find  $\mathbf{x}$  such that

$$A\mathbf{x} = \mathbf{y} \quad (1)$$

where  $A$  is an  $m \times n$  ( $m < n$ ) measurement matrix,  $\mathbf{y}$  is a  $m \times 1$  measurement vector and  $\mathbf{x}$  is an  $n \times 1$  unknown vector with only  $k$  ( $k < m$ ) nonzero components. We will further assume that the number of the measurements is  $m = \delta n$  and the number of the nonzero components of  $\mathbf{x}$  is  $k = \zeta n$ , where  $0 < \zeta < 1$  and  $0 < \delta < 1$  are constants independent of  $n$  (clearly,  $\delta > \zeta$ ).

A particular way of solving (1) which has recently generated a large amount of research is called  $\ell_1$ -optimization (basis pursuit) [2]. It proposes solving the following problem

$$\begin{aligned} \min \quad & \|\mathbf{x}\|_1 \\ \text{subject to} \quad & A\mathbf{x} = \mathbf{y}. \end{aligned} \quad (2)$$

Quite remarkably in [2] the authors were able to show that if the number of the measurements is  $m = \delta n$  and if the matrix  $A$  satisfies a special property called the restricted isometry property (RIP), then any unknown vector  $\mathbf{x}$  with no more than  $k = \zeta n$  (where  $\zeta$  is an absolute constant which is a function of  $\delta$ , but independent of  $n$ , and explicitly bounded in [2]) non-zero elements can be recovered by solving (2). Instead of characterizing the  $m \times n$  matrix  $A$  through the RIP condition, in [1], [3] the authors assume that  $A$  constitutes a  $k$ -neighborly polytope. It turns out (as shown in [1]) that this characterization of the matrix  $A$  is in fact a necessary and sufficient condition for (2) to produce the solution of (1). Furthermore, using the results of [4][7][8], it can be shown

that if the that if the matrix  $A$  has i.i.d. zero-mean Gaussian entries with overwhelming probability it also constitutes a  $k$ -neighborly polytope. The precise relation between  $m$  and  $k$  in order for this to happen is characterized in [1] as well.

In this paper we will be interested in providing new decoding algorithms which can increase the sparsity recovery thresholds and providing the theoretical guarantees for the emerging iterative reweighted  $\ell_1$  algorithms [16]. In [16], the algorithms iteratively updated weights for each element of  $\mathbf{x}$  in the objective function of  $\ell_1$  minimization, based on the decoding results from previous iterations. Experiments showed that the iterative reweighted  $\ell_1$  algorithms can greatly enhance the recoverable sparsity threshold for certain types of signals, for example, sparse signals with Gaussian entries. However, no rigorous theoretical results have been provided for establishing this phenomenon. To quote from [16], “any result quantifying the improvement of the reweighted algorithm for special classes of sparse or nearly sparse signals would be significant”. In this paper, we give simple reweighted  $\ell_1$  minimization algorithms for sparse signals recovery and try to provide a theoretical foundation for analyzing the iterative reweighted  $\ell_1$  algorithms. In particular, we show that for a nontrivial class of signals, (It is worth noting that empirically, the iterative reweighted  $\ell_1$  algorithms do not always improve the recoverable sparsity thresholds, for example, they often fail to improve the recoverable sparsity thresholds when the non-zero elements of the signals are “flat” [16]), a modified iterative reweighted  $\ell_1$  minimization algorithm can indeed deliver recoverable sparsity thresholds larger than those given in [1], [3] for unweighted  $\ell_1$  minimization algorithms. Our results are based on a high-dimensional geometrical analysis (Grassmann angle analysis) of the null-space characterization for  $\ell_1$  minimization and weighted  $\ell_1$  minimization algorithms. The main idea is to show that the preceding  $\ell_1$  minimization iterations can provide certain information about the support set of the signals and this support set information can be properly taken advantage of to perfectly recover the signals even though the sparsity of the signal  $\mathbf{x}$  itself is large.

This paper is structured as follows. In Section II, we present the iterative reweighted  $\ell_1$  algorithm for analysis. The signal model for  $\mathbf{x}$  will be given in Section III. In Section IV and Section V, we will show how the iterative reweighted  $\ell_1$  minimization algorithm can indeed improve recoverable sparsity thresholds. Numerical results will be given in Section

VI.

## II. THE MODIFIED ITERATIVE REWEIGHTED $\ell_1$ MINIMIZATION ALGORITHM

Let  $w_i^t, i = 1, \dots, n$ , denote the weights for the  $i$ -th element  $\mathbf{x}_i$  of  $\mathbf{x}$  in the  $t$ -th iteration of the iterative reweighted  $\ell_1$  minimization algorithm and let  $\mathbf{W}^t$  be the diagonal matrix with  $w_1^t, w_2^t, \dots, w_n^t$  on the diagonal. In the paper [16], the following iterative reweighted  $\ell_1$  minimization algorithm is presented:

*Algorithm 1:* [16]

- 1) Set the iteration count  $t$  to zero and  $w_i^t = 1, i = 1, \dots, n$ .
- 2) Solve the weighted  $\ell_1$  minimization problem

$$\mathbf{x}^t = \arg \min \|\mathbf{W}^t \mathbf{x}\|_1 \quad \text{subject to} \quad \mathbf{y} = \mathbf{A} \mathbf{x}. \quad (3)$$

- 3) Update the weights: for each  $i = 1, \dots, n$ ,

$$w_i^{t+1} = \frac{1}{|\mathbf{x}_i^t| + \epsilon'}, \quad (4)$$

where  $\epsilon'$  is a tunable positive number.

- 4) Terminate on convergence or when  $t$  attains a specified maximum number of iterations  $t_{\max}$ . Otherwise, increment  $t$  and go to step 2.

For the sake of tractable analysis, we will give another iterative reweighted  $\ell_1$  minimization algorithm, but it still captures the essence of the reweighted  $\ell_1$  algorithm presented in [16]. In our modified algorithm, we only do two  $\ell_1$  minimization programming, namely we stop at the time index  $t = 1$ .

*Algorithm 2:* 1) Set the iteration count  $t$  to zero and  $w_i^t = 1, i = 1, \dots, n$ .

- 2) Solve the weighted  $\ell_1$  minimization problem

$$\mathbf{x}^t = \arg \min \|\mathbf{W}^t \mathbf{x}\|_1 \quad \text{subject to} \quad \mathbf{y} = \mathbf{A} \mathbf{x}. \quad (5)$$

- 3) Update the weights: find the index set  $K' \subset \{1, 2, \dots, n\}$  which corresponds to the largest  $(1 - \epsilon)\rho_F(\delta)\delta n$  elements of  $\mathbf{x}^0$  in amplitudes, where  $0 < \epsilon < 1$  is a specified parameter and  $\rho_F(\delta)$  is the weak threshold for perfect recovery defined in [1] using  $\ell_1$  minimization (thus  $\zeta = \rho_F(\delta)\delta$  is the weak sparsity threshold). Then assign the weight  $W_1 = 1$  to those  $w_i^{t+1}$  corresponding to the set  $K'$  and assign the weight  $W_2 = W, W > 1$ , to those  $w_i^{t+1}$  corresponding to the complementary set  $\bar{K}' = \{1, 2, \dots, n\} \setminus K'$ .
- 4) Terminate on convergence or when  $t = 1$ . Otherwise, increment  $t$  and go to step 2.

This modified algorithm is certainly different from the algorithm from [16], but the important thing is that both algorithms assign bigger weights to those elements of  $\mathbf{x}$  which are more likely to be 0.

## III. SIGNAL MODEL FOR $\mathbf{x}$

In this paper, we consider the following model for the  $n$ -dimensional sparse signal  $\mathbf{x}$ . First of all, we assume that there exists a set  $K \subset \{1, 2, \dots, n\}$  with cardinality  $|K| = (1 - \epsilon)\rho_F(\delta)\delta n$  such that each of the elements of  $\mathbf{x}$  over the set  $K$  is large in amplitude. W.L.O.G., those elements are assumed to

be all larger than  $a_1 > 0$ . For a given signal  $\mathbf{x}$ , one might take such set  $K$  to be the set corresponding to the  $(1 - \epsilon)\rho_F(\delta)\delta n$  largest elements of  $\mathbf{x}$  in amplitude.

Secondly, (let  $\bar{K} = \{1, 2, \dots, n\} \setminus K$ ), we assume that the  $\ell_1$  norm of  $\mathbf{x}$  over the set  $\bar{K}$ , denoted by  $\|\mathbf{x}_{\bar{K}}\|_1$ , is upper-bounded by  $\Delta$ , though  $\Delta$  is allowed to take a non-diminishing portion of the total  $\ell_1$  norm  $\|\mathbf{x}\|_1$  as  $n \rightarrow \infty$ . We further denote the support set of  $\mathbf{x}$  as  $K_{\text{total}}$  and its complement as  $\bar{K}_{\text{total}}$ . The sparsity of the signal  $\mathbf{x}$ , namely the total number of nonzero elements in the signal  $\mathbf{x}$  is then  $|K_{\text{total}}| = k_{\text{total}} = \xi n$ , where  $\xi$  can be above the weak sparsity threshold  $\zeta = \rho_F(\delta)\delta$  achievable using the  $\ell_1$  algorithm.

In the following sections, we will show that if certain conditions on  $a_1$ ,  $\Delta$  and the measurement matrix  $\mathbf{A}$  are satisfied, we will be able to recover perfectly the signal  $\mathbf{x}$  using Algorithm 2 even though its sparsity level is above the sparsity threshold for  $\ell_1$  minimization. Intuitively, this is because the weighted  $\ell_1$  minimization puts larger weights on the signal elements which are more likely to be zero, and puts smaller weights on the signal support set, thus promoting sparsity at the right positions. In order to achieve this, we need some prior information about the support set of  $\mathbf{x}$ , which can be obtained from the decoding results in previous iterations. We will first argue that the equal-weighted  $\ell_1$  minimization of Algorithm 2 can sometimes provide very good information about the support set of signal  $\mathbf{x}$ .

## IV. ESTIMATING THE SUPPORT SET FROM THE $\ell_1$ MINIMIZATION

Since the set  $K'$  corresponds to the largest elements in the decoding results of  $\ell_1$  minimization, one might guess that most of the elements in  $K'$  are also in the support set  $K_{\text{total}}$ . The goal of this section is to get an upper bound on the cardinality of the set  $\bar{K}_{\text{total}} \cap K'$ , namely the number of zero elements of  $\mathbf{x}$  over the set  $K'$ . To this end, we will first give the notion of “weak” robustness for the  $\ell_1$  minimization.

Let  $K$  be fixed and  $\mathbf{x}_K$ , the value of  $\mathbf{x}$  on this set, be also fixed. Then the solution produced by (2),  $\hat{\mathbf{x}}$ , will be called *weakly robust* if, for some  $C > 1$  and all possible  $\mathbf{x}_{\bar{K}}$ , it holds that

$$\|(\mathbf{x} - \hat{\mathbf{x}})_{\bar{K}}\|_1 \leq \frac{2C}{C-1} \|\mathbf{x}_{\bar{K}}\|_1.$$

The above “weak” notion of robustness allows us to bound the error  $\|\mathbf{x} - \hat{\mathbf{x}}\|_1$  in the following way. If the matrix  $\mathbf{A}_K$ , obtained by retaining only those columns of  $\mathbf{A}$  that are indexed by  $K$ , has full column rank, then the quantity

$$\kappa = \max_{\mathbf{A}\mathbf{w}=0, \mathbf{w} \neq 0} \frac{\|\mathbf{w}_K\|_1}{\|\mathbf{w}_{\bar{K}}\|_1},$$

must be finite ( $\kappa < \infty$ ). In particular, since  $\mathbf{x} - \hat{\mathbf{x}}$  is in the null space of  $\mathbf{A}$  ( $\mathbf{y} = \mathbf{A}\mathbf{x} = \mathbf{A}\hat{\mathbf{x}}$ ), we have

$$\begin{aligned} \|\mathbf{x} - \hat{\mathbf{x}}\|_1 &= \|(\mathbf{x} - \hat{\mathbf{x}})_K\|_1 + \|(\mathbf{x} - \hat{\mathbf{x}})_{\bar{K}}\|_1 \\ &\leq (1 + \kappa) \|(\mathbf{x} - \hat{\mathbf{x}})_{\bar{K}}\|_1 \\ &\leq \frac{2C(1 + \kappa)}{C-1} \|\mathbf{x}_{\bar{K}}\|_1, \end{aligned}$$

thus bounding the recovery error. We can now give necessary and sufficient conditions on the measurement matrix  $A$  to satisfy the notion of weak robustness for  $\ell_1$  minimization.

*Theorem 1:* For a given  $C > 1$ , support set  $K$ , and  $\mathbf{x}_K$ , the solution  $\hat{\mathbf{x}}$  produced by (2) will be *weakly robust* if, and only if,  $\forall \mathbf{w} \in \mathbf{R}^n$  such that  $A\mathbf{w} = 0$ , we have

$$\|\mathbf{x}_K + \mathbf{w}_K\|_1 + \left\| \frac{\mathbf{w}_{\bar{K}}}{C} \right\|_1 \geq \|\mathbf{x}_K\|_1; \quad (6)$$

*Proof:* Sufficiency: Let  $\mathbf{w} = \hat{\mathbf{x}} - \mathbf{x}$ , for which  $A\mathbf{w} = A(\hat{\mathbf{x}} - \mathbf{x}) = 0$ . Since  $\hat{\mathbf{x}}$  is the minimum  $\ell_1$  norm solution, we have  $\|\mathbf{x}\|_1 \geq \|\hat{\mathbf{x}}\|_1 = \|\mathbf{x} + \mathbf{w}\|_1$ , and therefore  $\|\mathbf{x}_K\|_1 + \|\mathbf{x}_{\bar{K}}\|_1 \geq \|\hat{\mathbf{x}}_K\|_1 + \|\hat{\mathbf{x}}_{\bar{K}}\|_1$ . Thus,

$$\begin{aligned} \|\mathbf{x}_K\|_1 - \|\mathbf{x}_K + \mathbf{w}_K\|_1 &\geq \|\mathbf{w}_{\bar{K}} + \mathbf{x}_{\bar{K}}\|_1 - \|\mathbf{x}_{\bar{K}}\|_1 \\ &\geq \|\mathbf{w}_{\bar{K}}\|_1 - 2\|\mathbf{x}_{\bar{K}}\|_1. \end{aligned}$$

But the condition (6) guarantees that

$$\|\mathbf{w}_{\bar{K}}\|_1 \geq C(\|\mathbf{x}_K\|_1 - \|\mathbf{x}_K + \mathbf{w}_K\|_1),$$

so we have

$$\|\mathbf{w}_{\bar{K}}\|_1 \leq \frac{2C}{C-1} \|\mathbf{x}_{\bar{K}}\|_1,$$

and

$$\|\mathbf{x}_K\|_1 - \|\hat{\mathbf{x}}_K\|_1 \leq \frac{2}{C-1} \|\mathbf{x}_{\bar{K}}\|_1,$$

as desired.

*Necessity:* Since in the above proof of the sufficiency, equalities can be achieved in the triangular inequalities, the condition (6) is also a necessary condition for the weak robustness to hold for every  $\mathbf{x}$ . (Otherwise, for certain  $\mathbf{x}$ 's, there will be  $\mathbf{x}' = \mathbf{x} + \mathbf{w}$  with  $\|\mathbf{x}'\|_1 < \|\mathbf{x}\|_1$  while violating the respective robustness definitions. Also, such  $\mathbf{x}'$  can be the solution to (2)). ■

We should remark (without proof for the interest of space) that for any  $\delta > 0$ ,  $0 < \epsilon < 1$ , let  $|K| = (1 - \epsilon)\rho_F(\delta)\delta n$ , and suppose each element of the measurement matrix  $A$  is sampled from i.i.d. Gaussian distribution, then there exists a constant  $C > 1$  (as a function of  $\delta$  and  $\epsilon$ ), such that the condition (6) is satisfied with overwhelming probability as the problem dimension  $n \rightarrow \infty$ . At the same time, the parameter  $\kappa$  defined above is upper-bounded by a finite constant (independent of the problem dimension  $n$ ) with overwhelming probability as  $n \rightarrow \infty$ . These claims can be shown by using the Grassmann angle approach for the balancedness property of random linear subspaces in [12]. In the current version of our paper, we would make no attempt to explicitly express the parameters  $C$  and  $\kappa$ .

In Algorithm 2, after equal-weighted  $\ell_1$  minimization, we pick the set  $K'$  corresponding to the  $(1 - \epsilon)\rho_F(\delta)\delta$  largest elements in amplitudes from the decoding result  $\hat{\mathbf{x}}$  (namely  $\mathbf{x}^0$  in the algorithm description) and assign the weights  $W_1 = 1$  to the corresponding elements in the next iteration of reweighted  $\ell_1$  minimization. Now we can show that an overwhelming portion of the set  $K'$  are also in the support set  $K_{\text{total}}$  of  $\mathbf{x}$  if the measurement matrix  $A$  satisfies the specified weak robustness property.

*Theorem 2:* Supposed that we are given a signal vector  $\mathbf{x} \in \mathbf{R}^n$  satisfying the signal model defined in Section III. Given  $\delta > 0$ , and a measurement matrix  $A$  which satisfies the weak robustness condition in (6) with its corresponding  $C > 1$  and  $\kappa < \infty$ , then the set  $K'$  generated by the equal-weighted  $\ell_1$  minimization in Algorithm 2 contains at most  $\frac{2C}{(C-1)^{\frac{a_1}{2}}} \|\mathbf{x}_{\bar{K}}\|_1 + \frac{2C\kappa}{(C-1)^{\frac{a_1}{2}}} \|\mathbf{x}_K\|_1$  indices which are outside the support set of signal  $\mathbf{x}$ .

*Proof:* Since the measurement matrix  $A$  satisfies the weak robustness condition for the set  $K$  and the signal  $\mathbf{x}$ ,

$$\|(\mathbf{x} - \hat{\mathbf{x}})_{\bar{K}}\|_1 \leq \frac{2C}{C-1} \|\mathbf{x}_{\bar{K}}\|_1.$$

By the definition of the  $\kappa < \infty$ , namely,

$$\kappa = \max_{A\mathbf{w}=0, \mathbf{w} \neq 0} \frac{\|\mathbf{w}_K\|_1}{\|\mathbf{w}_{\bar{K}}\|_1},$$

we have

$$\|(\mathbf{x} - \hat{\mathbf{x}})_K\|_1 \leq \kappa \|(\mathbf{x} - \hat{\mathbf{x}})_{\bar{K}}\|_1.$$

Then there are at most  $\frac{2C}{(C-1)^{\frac{a_1}{2}}} \|\mathbf{x}_{\bar{K}}\|_1$  indices that are outside the support set of  $\mathbf{x}$  but have amplitudes larger than  $\frac{a_1}{2}$  in the corresponding positions of the decoding result  $\hat{\mathbf{x}}$  from the equal-weighted  $\ell_1$  minimization algorithm. This bound follows easily from the facts that all such indices are in the set  $\bar{K}$  and that  $\|(\mathbf{x} - \hat{\mathbf{x}})_{\bar{K}}\|_1 \leq \frac{2C}{C-1} \|\mathbf{x}_{\bar{K}}\|_1$ .

Similarly, there are at most  $\frac{2C\kappa}{(C-1)^{\frac{a_1}{2}}} \|\mathbf{x}_{\bar{K}}\|_1$  indices which are originally in the set  $K$  but now have corresponding amplitudes smaller than  $\frac{a_1}{2}$  in the decoded result  $\hat{\mathbf{x}}$  of the equal-weighted  $\ell_1$  algorithm.

Since the set  $K'$  corresponds to the largest  $(1 - \epsilon)\rho_F(\delta)\delta n$  elements of the signal  $\hat{\mathbf{x}}$ , by combining the previous two results, it is not hard to see that the number of indices which are outside the support set of  $\mathbf{x}$  but are in the set  $K'$  is no bigger than  $\frac{2C}{(C-1)^{\frac{a_1}{2}}} \|\mathbf{x}_{\bar{K}}\|_1 + \frac{2C\kappa}{(C-1)^{\frac{a_1}{2}}} \|\mathbf{x}_{\bar{K}}\|_1$ . ■

As we can see, Theorem 2 provides useful information about the support set of the signal  $\mathbf{x}$ , which can be used in the analysis for the weighted  $\ell_1$  minimization using the null-space Grassmann Angle analysis approach for weighted  $\ell_1$  minimization algorithm [13].

## V. THE GRASSMANN ANGLE APPROACH FOR THE REWEIGHTED $\ell_1$ MINIMIZATION

In the previous work [13], the authors have shown that by exploiting certain prior information about the original signal, it is possible to extend the threshold of sparsity factor for successful recovery beyond the original bounds of [1], [3]. The authors proposed a nonuniform sparsity model in which the entries of the vector  $\mathbf{x}$  can be considered as  $T$  different classes, where in the  $i$ th class, each entry is (independently from others) nonzero with probability  $P_i$ , and zero with probability  $1 - P_i$ . The signals generated based on this model will have around  $n_1 P_1 + \dots + n_T P_T$  nonzero entries with high probability, where  $n_i$  is the size of the  $i$ th class. Examples of such signals arise in many applications as medical or natural

imaging, satellite imaging, DNA micro-arrays, network monitoring and so on. They prove that provided such structural prior information is available about the signal, a proper *weighted*  $\ell_1$ -minimization strictly outperforms the regular  $\ell_1$ -minimization in recovering signals with some fixed average sparsity from under-determined linear i.i.d. Gaussian measurements.

The detailed analysis in [13] is only done for  $T = 2$ , and is based on the high dimensional geometrical interpretations of the constrained weighted  $\ell_1$ -minimization problem:

$$\min_{\mathbf{Ax}=\mathbf{y}} \sum_{i=1}^n w_i |\mathbf{x}_i| \quad (7)$$

Let the two classes of entries be denoted by  $K_1$  and  $K_2$ . Also, due to the partial symmetry, for any suboptimal set of weights  $\{w_1, \dots, w_n\}$  we have the following

$$\forall i \in \{1, 2, \dots, n\} \quad w_i = \begin{cases} W_1 & \text{if } i \in K_1 \\ W_2 & \text{if } i \in K_2 \end{cases}$$

The following theorem is implicitly proven in [13] and more explicitly stated and proven in [14]

**Theorem 3:** Let  $\gamma_1 = \frac{n_1}{n}$  and  $\gamma_2 = \frac{n_2}{n}$ . If  $\gamma_1, \gamma_2, P_1, P_2, W_1$  and  $W_2$  are fixed, there exists a critical threshold  $\delta_c = \delta_c(\gamma_1, \gamma_2, P_1, P_2, \frac{W_2}{W_1})$ , totally computable, such that if  $\delta = \frac{m}{n} \geq \delta_c$ , then a vector  $\mathbf{x}$  generated randomly based on the described nonuniformly sparse model can be recovered from the weighted  $\ell_1$ -minimization of (7) with probability  $1 - o(e^{-cn})$  for some positive constant  $c$ .

In [13] and [14], a way for computing  $\delta_c$  is presented which, in the uniform sparse case (e.g  $\gamma_2 = 0$ ) and equal weights, is consistent with the weak threshold of Donoho and Tanner for almost sure recovery of sparse signals with  $\ell_1$ -minimization.

In summary, given a certain  $\delta$ , the two different weights  $W_1$  and  $W_2$  for weighted  $\ell_1$  minimization, the size of the two weighted blocks, and also the number (or proportion) of nonzero elements inside each weighted block, the results from [13] can determine whether a uniform random measurement matrix will be able to perfectly recover the original signals with overwhelming probability. Using this framework we can now begin to analyze the performance of the modified re-weighted algorithm of section II. Although we are not directly given some prior information, as in the nonuniform sparse model for instance, about the signal structure, one might hope to infer such information after the first step of the modified re-weighted algorithm. To this end, note that the immediate step in the algorithm after the regular  $\ell_1$ -minimization is to choose the largest  $(1 - \epsilon)\rho_F(\delta)\delta n$  entries in absolute value. This is equivalent to splitting the index set of the vector  $\mathbf{x}$  to two classes  $K'$  and  $K''$ , where  $K'$  corresponds to the larger entries. We now try to find a correspondence between this setup and the setup of [13] where sparsity factors on the sets  $K'$  and  $\bar{K}'$  are known. We claim the following upper bound on the number of nonzero entries of  $\mathbf{x}$  with index on  $K'$ .

**Theorem 4:** There are at least  $(1 - \epsilon)\rho_F(\delta)\delta n - \frac{4C(\kappa+1)\Delta}{(C-1)a_1}$  nonzero entries in  $\mathbf{x}$  with index on the set  $K'$ .

*Proof:* Directly from Theorem 2 and the fact that  $\|\mathbf{x}_{\bar{K}}\|_1 \leq \Delta$ . ■

The above result simply gives us a lower bound on the sparsity factor (ratio of nonzero elements) in the vector  $\mathbf{x}_{K'}$

$$P_1 \geq 1 - \frac{4C(\kappa+1)}{(C-1)a_1\rho_F(\delta)\delta} \frac{\Delta}{n} \quad (8)$$

Since we also know the original sparsity of the signal,  $\|\mathbf{x}\|_0 \leq k_{\text{total}}$ , we have the following upper bound on the sparsity factor of the second block of the signal  $\mathbf{x}_{\bar{K}'}$

$$P_2 \leq \frac{k_{\text{total}} - (1 - \epsilon)\rho_F(\delta)\delta n + \frac{4C(\kappa+1)\Delta}{(C-1)a_1}}{n - (1 - \epsilon)\rho_F(\delta)\delta n} \quad (9)$$

Note that if  $a_1$  is large and  $1 \gg \frac{\Delta}{a_1 n}$  (Note however, we can let  $\Delta$  take a non-diminishing portion of  $\|\mathbf{x}\|_1$ , even though that portion can be very small), then  $P_1$  is very close to 1. This means that the original signal is much denser in the block  $K'$  than in the second block  $\bar{K}'$ . Therefore, as in the last step of the modified re-weighted algorithm, we may assign a weight  $W_1 = 1$  to all entries of  $\mathbf{x}$  in  $K'$  and weight  $W_2 = W$ ,  $W > 1$  to the entries of  $\mathbf{x}$  in  $\bar{K}'$  and perform the weighted  $\ell_1$ -minimization. The theoretical results of [13], namely Theorem 3 guarantee that as long as  $\delta > \delta_c(\gamma_1, \gamma_2, P_1, P_2, \frac{W_2}{W_1})$  then the signal will be recovered with overwhelming probability for large  $n$ .<sup>1</sup> The numerical examples in the next Section do show that the reweighted  $\ell_1$  algorithm can increase the recoverable sparsity threshold, i.e.  $P_1\gamma_1 + P_2\gamma_2$ .

## VI. NUMERICAL COMPUTATIONS ON THE BOUNDS

In this section, we demonstrate the numerical performance of the new reweighted algorithm in this paper. Let  $\delta = 0.555$  and  $\frac{W_2}{W_1} = 3$  be fixed, meaning that  $\zeta = \rho_F(\delta)\delta$  is also given. We set  $\epsilon = 0.01$ . The sizes of the two classes  $K'$  and  $\bar{K}'$  would then be  $\gamma_1 n = (1 - \epsilon)\zeta n$  and  $\gamma_2 n = (1 - \gamma_1)n$  respectively. Here we take  $n = 200$  and for each sparsity  $k_{\text{total}}$ , we perform 250 experiments with randomly generated i.i.d. Gaussian measurement matrices and randomly generated signals from certain distributions.

The two red curves are for the recovery Gaussian and uniform  $[-1, 1]$  distributed signals using the proposed reweighted algorithm, and we observed that they basically have the same behavior: coincidentally the two distributions are both nonzero at the origin. The black curve is for a chi-squared distribution with 2 degrees of freedom (its distribution is zero at the origin, but the derivative is nonzero) and has a worse threshold. The purple curve is a chi-squared with 4 degrees of freedom (the distribution and its first derivative at the origin are zero, but not the second derivative) and the light blue curve is a chi-squared with 6 degrees of freedom (here even the second derivative is

<sup>1</sup>We should remark that this only holds if the Gaussian random matrix is sampled independently from the signal to be decoded in the weighted  $\ell_1$  minimization. In the iterative reweighted  $\ell_1$  minimization, we do not have this independence. However, this can be accounted for by using a union bound over the possible configurations of the set  $K'$ . Using similar arguments as in Theorem 2, we can show that the exponent for this union bound can be made arbitrarily small if  $1 \gg \frac{\Delta}{a_1 n}$ , which can be outweighed by the Grassmann Angle exponent. Due to space limitations, we will leave the proofs on these arguments for a full version of this paper.

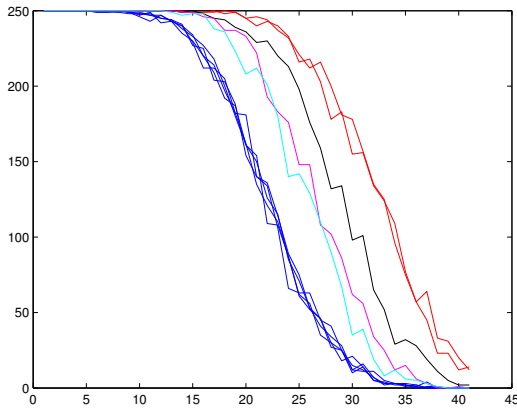


Fig. 1. Sparsity  $k_{total}$  (the horizontal axis) and Perfect Recovery Frequency (the vertical axis)

zero at the origin). As you can see, the more derivatives that vanish at the origin, the worse the threshold.

There are also 5 blue curves, one for each of the above cases, and they show the performance of the  $\ell_1$  optimization. As you can see, they all behave the same, since  $\ell_1$  optimization cannot exploit the distribution of the nonzero entries.

The important thing to note is that our algorithm works exceedingly well. The problem at hand is one for which  $m/n = 0.55$ , the reweighted algorithm has a threshold of around  $k_{total} = 33$  for  $n = 200$ , which is a 50 percent improvement over the  $\ell_1$  minimization method.

## REFERENCES

- [1] David Donoho, "High-Dimensional Centrally Symmetric Polytopes with Neighborliness Proportional to Dimension", *Discrete and Computational Geometry*, 102(27), pp. 617-652, 2006, Springer.
- [2] Emmanuel Candes and Terence Tao, "Decoding by linear programming", *IEEE Trans. on Information Theory*, 51(12), pp. 4203 - 4215, December 2005.
- [3] David Donoho and Jared Tanner, "Thresholds for the Recovery of Sparse Solutions via L1 Minimization", *Proceedings of the Conference on Information Sciences and Systems*, March 2006.
- [4] A.M. Vershik and P.V. Sporyshev, "Asymptotic Behavior of the Number of Faces of Random Polyhedra and the Neighborliness Problem", *Selecta Mathematica Sovietica* vol. 11, No. 2 (1992).
- [5] Branko Grünbaum, Grassmann angles of convex polytopes. *Acta Math.*, 121:pp.293-302, 1968.
- [6] Branko Grünbaum, *Convex polytopes*, volume 221 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 2003. Prepared and with a preface by Volker Kaibel, Victor Klee and Günter M. Ziegler.
- [7] Fernando Affentranger and Rolf Schneider. "Random projections of regular simplices". *Discrete Comput. Geom.*, 7(3):pp.219-226, 1992.
- [8] Károly Böröczky, Jr. and Martin Henk. "Random projections of regular polytopes". *Arch. Math. (Basel)*, 73(6):pp.465-473, 1999.
- [9] L.A. Santaló, Geometría integral en espacios de curvatura constante, Rep. Argentina Publ. Com. Nac. Energ. Atómica, Ser. Mat. 1, No. 1 (1952)
- [10] Peter McMullen. "Non-linear angle-sum relations for polyhedral cones and polytopes". *Math. Proc. Cambridge Philos. Soc.*, 78(2):pp.247-261, 1975.
- [11] Emmanuel Candes, "Compressive sampling", *Proc. International Congress of Mathematics*, 3, pp. 1433-1452, Madrid, Spain, 2006.
- [12] Weiyu Xu and Babak Hassibi, "Compressive Sensing over the Grassmann Manifold: A Unified Analytical Approach," *Allerton Conference on Communication, Control, and Computing*, September 2008

- [13] M. Amin Khajehnejad, Weiyu Xu, Amir Salman Avestimehr, Babak Hassibi, "Weighted  $\ell_1$  Minimization for Sparse Recovery with Prior Information", accepted to *the International Symposium on Information Theory 2009*, available online <http://arxiv.org/abs/0901.2912>
- [14] M. Amin Khajehnejad, Weiyu Xu, Amir Salman Avestimehr, Babak Hassibi, "Weighted  $\ell_1$  Minimization for Sparse Recovery with Prior Information", *journal version, to be submitted*
- [15] Weiyu Xu and Babak Hassibi, "On Sharp Performance Bounds for Robust Sparse Signal Recoveries", accepted to *the International Symposium on Information Theory 2009*
- [16] E. J. Candes, M. B. Wakin, and S. Boyd, "Enhancing Sparsity by Reweighted  $\ell_1$  Minimization", *Journal of Fourier Analysis and Applications*, 14(5), pp. 877-905, special issue on sparsity, December 2008